A novel gene, *NSD1*, is fused to *NUP98* in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia

Rina J. Jaju, Carrie Fidler, Oskar A. Haas, Amanda J. Strickson, Fiona Watkins, Kevin Clark, Nicholas C. P. Cross, Jan-Fang Cheng, Peter D. Aplan, Lyndal Kearney, Jacqueline Boultwood, and James S. Wainscoat

The recurrent translocation t(5;11)(q35; p15.5) associated with a 5q deletion, del(5q), has been reported in childhood acute myeloid leukemia (AML). We report the cloning of the translocation breakpoints in de novo childhood AML harboring a cryptic t(5;11)(q35;p15.5). Fluorescence in situ hybridization (FISH) analysis demonstrated that the nucleoporin gene (*NUP98*) at 11p15.5 was disrupted by this translocation. By using 3′–rapid amplification of complementary DNA ends (3′-RACE) polymerase chain reaction, we identified a chimeric messenger RNA that results in the in-frame fusion of *NUP98* to a novel gene, *NSD1.* The *NSD1* gene has 2596 amino acid residues and a 85% homology to the murine *Nsd1* with the domain structure being conserved. The *NSD1* gene was localized to 5q35 by FISH and is widely expressed. The reciprocal transcript, *NSD1-NUP98*, was also detected by reverse transcriptase–polymerase chain reaction. This is the first report in which the novel gene *NSD1* has been implicated in human malignancy. (Blood. 2001;98:1264-1267)

Introduction

Recurring chromosome translocations are common in a wide spectrum of hematologic malignancies.[1-3] We have reported the identification of a recurrent cryptic translocation t(5;11)(q35; p15.5) associated with a deletion of the long arm of chromosome 5, del(5q), in de novo childhood acute myeloid leukemia (AML).[4] Using fluorescence in situ hybridization (FISH), we localized the breakpoint on 11p15.5 between Harvey ras-1 related gene complex (*HRC*) and radixin pseudogene *(RDPX1).* The nucleoporin 98 gene (*NUP98*) lies inside this region, and its rearrangement has been documented in both de novo and therapy-related myelodysplastic syndrome, AML, and acute lymphoblastic leukemia in children and adults.[5-13] We speculated that *NUP98* was a good candidate gene for the involvement in the t(5;11) AML.

We now confirm that the chromosome 11 breakpoint gene is *NUP98*, and we report the cloning of its novel fusion partner, nuclear receptor-binding Su(var), Enhancer of zeste [E(z)], and Trithorax (Trx) (SET) domain protein 1 (*NSD1*), as the chromosome 5 breakpoint gene in the recurrent t(5;11)(q35;p15.5) in childhood AML.

Study design

Donor samples

We analyzed samples from 2 patients with childhood AML-M2 with t(5;11)(q35;p15.5) and a del(5q). The clinical and cytogenetic data of these 2 cases have been previously reported as patients nos. 1 and 3.[4] Detailed molecular analysis was performed on the leukemic sample of patient no. 3 because RNA was available only from this patient.

FISH analysis

For FISH studies PAC 1173K1 (containing exons 10-20 of the *NUP98* gene), plasmids p9R1 (containing exons 10-12 of the *NUP98* gene),[14] and p6G2 (containing exons 13-14 of the *NUP98* gene)[14] from 11p15.5 were used. A 5p probe, cos113-1,[15] was used to identify both chromosomes 5. A 1.2-kilobase (kb) complementary DNA (cDNA) clone of the human *NSD1* (IMAGE clone 2621298) was also used as a probe for FISH studies. FISH analysis was performed on bone marrow metaphases as previously described.[4]

Nucleic acid isolation

Total RNA was extracted from cryopreserved leukemic cell suspensions using the Totally RNA kit (Ambion, Austin, TX). Plasmid DNA was extracted using Qiagen reagents and protocols (Qiagen miniprep kit, Qiagen, Crawley, United Kingdom).

3′-RACE

The 3′-RACE (3′–rapid amplification of cDNA ends) was performed using the SMART-RACE cDNA amplification kit and protocol (Clontech, Palo Alto, CA). Briefly, first-strand cDNA was reverse transcribed from 1 μg total RNA using Superscript II and the 3′-RACE cDNA synthesis primer (3′-CDS) from the kit. An aliquot of the first-strand cDNA was then amplified using a *NUP98* gene–specific forward primer (NUP98-1, Table 1) and a universal primer mix (SMART-RACE kit). Polymerase chain reaction (PCR) conditions were as described by the manufacturer. A nested PCR reaction using the nested universal primer (SMART-RACE kit) as the reverse primer and NUP98-2 (internal to NUP98-1, Table 1) as the forward

From the LRF Molecular Haematology Unit, Nuffield Department of Clinical Laboratory Sciences, University of Oxford; and MRC Molecular Haematology Unit, Institute of Molecular Medicine; both of John Radcliffe Hospital, Oxford, United Kingdom; Children's Cancer Research Institute, St Anna Kinderspital, Vienna, Austria; Department of Haematology, Imperial College School of Medicine, Hammersmith Hospital, London, United Kingdom; Life Sciences Division, Lawrence Berkeley National Laboratory, CA; and Division of Clinical Sciences, National Cancer Institute, Gaithersburg, MD.

Submitted January 16, 2001; accepted April 17, 2001.

Supported by the Leukaemia Research Fund, United Kingdom (R.J.J., C.F., A.J.S., F.W., J.B., J.S.W.), and the Medical Research Council (K.C, L.K).

Reprints: Rina J. Jaju, LRF Molecular Haematology Unit, Nuffield Dept of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom; e-mail: rina.jaju@ndcls.ox.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

Table 1. Oligonucleotide primer sequences

Primer	Oligonucleotide sequence (5′ → 3′)	Gene	Nucleotides
NUP98-1	GGGACTCTTGGAACTGGGCTTGGTGCAG	*NUP98*	1372-1399
NUP98-2	GGAGGGCCTCTTGGTACAGGAGCCTTTGG	*NUP98*	1462-1490
NUP98-3	GTGCTGGACAGGCATCTTTG	*NUP98*	1418-1437
NUP98-4	TCCAAAGCCCAAAGTGGCTG	*NUP98*	1533-1514
NUP98-5	TCTTGGTACAGGAGCCTTTG	*NUP98*	1470-1489
NSD1-1	TCCAAAAGCCACTTGCTTGGC	*NSD1*	3557-3537
NSD1-2	GGGCCACGGTTAAATGTTTG	*NSD1*	3400-3419
NUP98-6	GACATCGGATTCCGGAAGAG	*NUP98*	1664-1644

Nucleotide coordinates refer to GenBank accession numbers U41815 (*NUP98*) and AF322907 (*NSD1*).

primer was performed according to the manufacturer's instructions. Second-round PCR products were electrophoresed, purified, and subcloned. Colonies with recombinant plasmids containing the PCR products were screened by hybridization using standard protocols.[16] An oligonucleotide probe of 116 base pairs (bp) (NUP-ex12) generated using primers NUP98-3 and NUP98-4 was used for screening (Table 1). All positive plasmid clones were selected for sequencing.

Reverse transcriptase–PCR

Reverse transcriptase (RT)-PCR was performed using the RT one-step RT-PCR kit (ABgene, Surrey, United Kingdom) with sense NUP98-5 and antisense NSD1-1 primers (Table 1). The PCR thermal cycling protocol was performed according to the manufacturer's instructions with an annealing temperature of 58°C. Similarly, RT-PCR was performed using sense NSD1-2 and antisense NUP98-6 primers and the same PCR conditions as above (Table 1). The RT-PCR products of both reactions were subcloned and sequenced.

Sequence analysis

Plasmid clones were sequenced using the Autoread Cy-5 sequencing kit and ALF express automated sequencer (Amersham-Pharmacia Biotech, Amersham, United Kingdom). Sequence analysis was performed using the Wisconsin Package Version 10.1 (Genetics Computer Group, Madison, WI).

Northern blot analysis

A cDNA clone (Image clone 2621298) of the human *NSD1* gene was used to probe human multiple tissue Northern blots according to the manufacturer's instructions (Clontech).

Results and discussion

In the current study, we identified *NUP98* as the chromosome 11 breakpoint gene by FISH using a PAC clone 1173K1 in 2 t(5;11)(q35;p15.5) AML cases (patients nos. 1 and 3) (Figure 1A). Further, FISH analyses using plasmids specific for exons of *NUP98* (p6G2 and p9R1) confirmed that the breakpoint was within *NUP98*, in the intron between exons 12 and 13 of this gene in patients nos. 1 and 3. Previous studies have shown the principal transcript in other *NUP98* translocations fuses the 5′ end of the *NUP98* gene in-frame to the 3′ end of the partner gene.[7-13] The 3′-RACE was performed on RNA isolated from patient no. 3 to identify the fusion partner of *NUP98*.

RACE-PCR products were purified and subcloned into plasmid vectors. A total of 88 recombinant clones were screened by hybridization and 45 clones hybridized to an internal *NUP98* oligonucleotide, NUP-ex12. Sequence analysis revealed that 13 of the 3′-RACE clones diverged from the germline *NUP98* sequence at nucleotide 1552 (GenBank accession number U41815). A BLAST search showed that the sequence immediately 3′ of this divergence was a novel human gene with 85% homology to the gene encoding murine *Nsd1* (GenBank accession number NM_008739).

The full coding sequence of this novel human gene was determined by the sequencing of cDNA and genomic clones. The sequence had an open reading frame of 7788 nucleotides and a predicted protein of 2596 amino acids (GenBank accession number AF322907). The coding sequence of the novel gene was compared with the coding sequence of murine *Nsd1*,[17] and 85% homology was found at the amino acid level. Therefore, this indicated that the gene was the human homolog of the mouse *Nsd1* gene, and it was designated *NSD1*. Sequence analysis of 3′-RACE products showed that *NUP98* and *NSD1* messenger RNA (mRNA) were fused in-frame joining nucleotide 1552 of *NUP98* to the nucleotide 3504 of *NSD1* (Figure 2A). The domain structure of the human *NSD1* gene product was the same as that of murine Nsd1 protein,[17] with conserved SET domain, SET domain–associated cysteine-rich domain (SAC domain), and 5 PHD fingers.

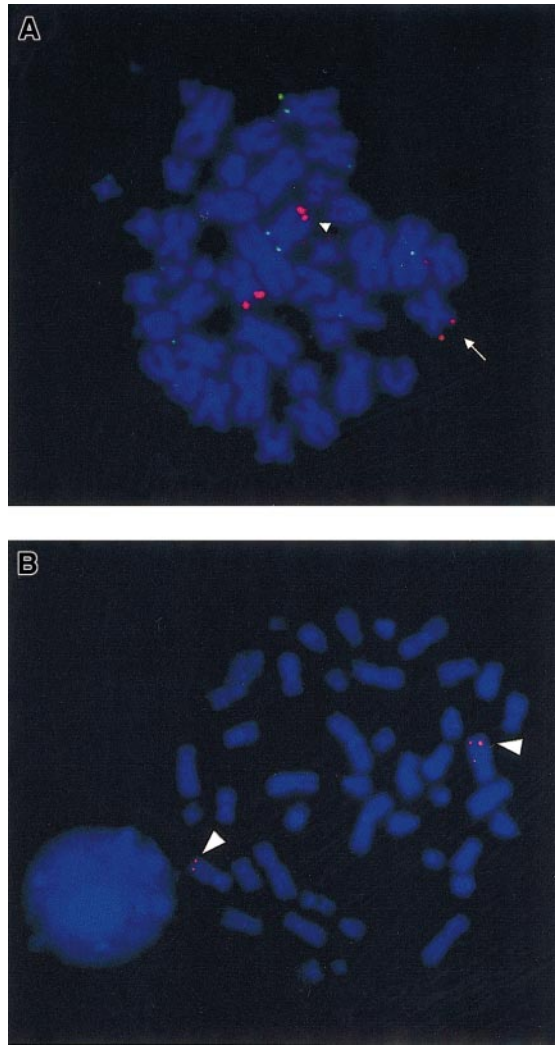

Figure 1. FISH analysis of *NUP98* rearrangement and chromosomal mapping of human *NSD1* gene. (A) FISH using PAC 1173K1 (containing *NUP98* gene) shows that this probe hybridizes to the normal chromosome 11, der(11) (arrow) and the der(5) (arrowhead), seen as red fluorescent signals in bone marrow metaphase from patient no. 3. The probe showing green fluorescent signals is the control 5p probe (cos113-1) used to identify both chromosomes 5. The same FISH result was also seen in metaphases from patient no. 1. (B) FISH using a cDNA clone of human *NSD1* (Image clone 2621298) as a probe on normal peripheral blood metaphase shows that this novel gene localizes to 5q35 (red fluorescent signals) (arrowhead).

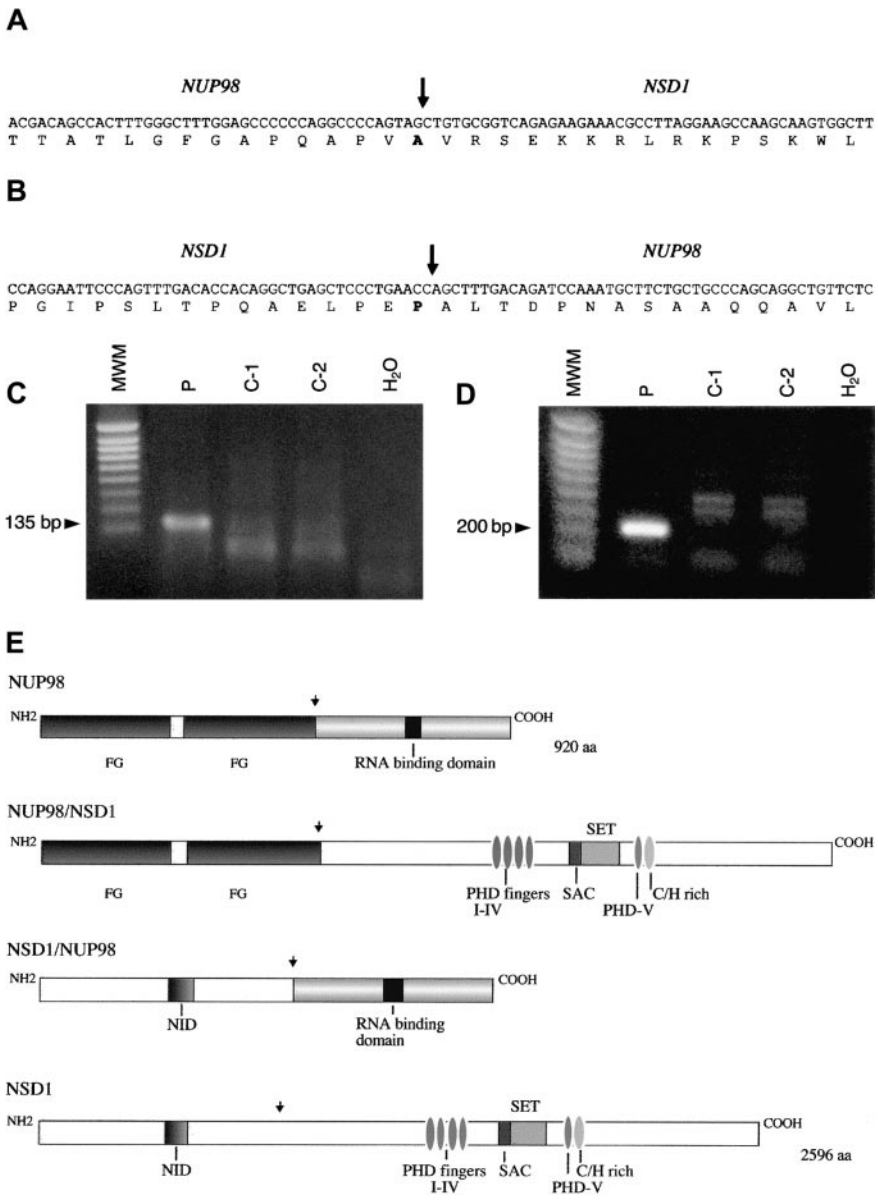

Figure 2. *NUP98-NSD1* and *NSD1-NUP98* fusion cDNA. Nucleotide and amino acid sequences around the junction of (A) *NUP98-NSD1* fusion transcript and (B) *NSD1-NUP98* fusion transcript. The *NUP98* and *NSD1* fusion points are shown by a vertical arrow. (C) RT-PCR assay for *NUP98-NSD1* fusion mRNA. Primers NUP98-5 and NSD1-1 were used to amplify the 135 bp fusion mRNA. (D) RT-PCR assay for *NSD1-NUP98* fusion mRNA. Primers NSD1-2 and NUP98-6 were used to amplify the 200 bp fusion mRNA. No specific products of the expected size are seen in the normal controls (C1-C2) and negative controls (H_2O) in panels C and D. MWM indicates 1 kb ladder molecular weight marker. (E) Schematic representations of the wild-type NUP98, NSD1, and fusion NUP98-NSD1, NSD1-NUP98 proteins. Functional domains indicated for NUP98 are FG repeat regions at the N-terminus and RNA binding domain at the C-terminus. Domains for NSD1 are N-terminus, nuclear receptor interaction domain (NID), $PHD_{I\text{-}IV}$ fingers, SET and SAC domains, PHD_V finger, and cysteine-histidine (C/H)-rich motif, and C-terminus. Functional domains of NSD1 are similar to the murine Nsd1 as designated in reference 17. NUP98 and NSD1 fusion points are shown by a vertical arrow.

RT-PCR on material from patient no. 3 was performed using primers flanking the *NUP98-NSD1* junction and gave a product of expected size (135 bp) (Figure 2C). Sequence analysis of RT-PCR products confirmed that *NUP98* and *NSD1* mRNA were fused in-frame at the same *NUP98* exon as reported in the literature.[5-8,10,11,13] A reciprocal fusion mRNA transcript *NSD1-NUP98* was also detected by RT-PCR analysis with an expected product size of 200 bp (Figure 2D). Sequence analysis showed that the mRNA was fused in-frame joining nucleotides 3503 of *NSD1* to nucleotide 1553 of *NUP98* (Figure 2B). Expression of both *NUP98-NSD1* and *NSD1-NUP98* transcripts suggests that both may have a biological role. The expression pattern of human *NSD1* in normal human tissues was analyzed using the novel cDNA clone as a probe and showed that this gene is widely expressed in hematologic and other tissues, and there are 2 transcripts of approximately 10.5 and 12 kb, respectively. FISH analysis using the novel cDNA clone revealed that the human *NSD1* was localized to 5q35 as expected (Figure 1B).

The *NUP98* gene encodes a 98-kd component of the nuclear pore complex and localizes to the nucleoplasmic side of the nuclear pore complex.[18] It is thought to function as a docking protein through the N-terminal domain of the protein, which contains the conserved multiple phenylalanine-glycine (FG) repeats.[18] The FG repeats of NUP98 have been shown to be retained in the various fusion transcripts.[5-13] The FG repeats were also retained in the NUP98-NSD1 fusion transcript in the patient reported here.

The fusion partner of *NUP98* in the t(5;11) AML is *NSD1*, which is reported here for the first time and is the human homolog of murine *Nsd1*. Murine Nsd1 contains 2 distinct nuclear receptor interaction domains and also contains distinct activation and repression domains, and it has been suggested that it may define a novel class of bifunctional transcriptional intermediary factors.[17] By comparison to the murine Nsd1 protein, the human NSD1 protein has the domain structure conserved and these domains are found in proteins involved in the epigenetic control of transcription; it therefore has been suggested that Nsd1 is involved in some aspects of transcriptional control at the chromatin level.[17] It is thought that the PHD domain is a protein-protein interaction domain and that SET/SAC domains at least in one protein serve as

a histone methyltransferase.[19-20] The schematic representations of NUP98-NSD1 and NSD1-NUP98 fusion proteins are shown in Figure 2E.

In summary, we have cloned the recurrent t(5;11) chromosomal translocation, which results in the previously undescribed in-frame fusion between *NUP98* and a novel gene *NSD1* in de novo childhood AML. Future studies based on the structure-function relationship of the *NSD1* gene and functional analysis of the 2 predicted chimeric proteins will provide an insight into the mechanism of leukemogenesis.

Acknowledgment

The authors thank Robin Roberts-Gant for assistance with the figure preparations.

References

1. Rabbitts TH. Chromosomal translocations in human cancer. Nature. 1994;372:143-149.
2. Look AT. Oncogenic transcription factors in the human acute leukemias. Science. 1997;278: 1059-64.
3. Solomon E, Borrow J, Goddard AD. Chromosome aberrations and cancer. Science. 1991;254:1153-1160.
4. Jaju RJ, Haas OA, Neat M, et al. A new recurrent translocation, t(5;11)(q35;p15.5), associated with del(5q) in childhood acute myeloid leukemia. Blood. 1999;94:773-780.
5. Borrow J, Shearman AM, Stanton VP Jr, et al. The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin *NUP98* and class I homeoprotein *HOXA9.* Nat Genet. 1996;12:159-167.
6. Nakamura T, Largaespada DA, Lee MP, et al. Fusion of the nucleoporin gene *NUP98* to *HOXA9* by the chromosome translocation t(7;11)(p15; p15) in human myeloid leukaemia. Nat Genet. 1996;12:154-158.
7. Arai Y, Hosoda F, Kobayashi H, et al. The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, *NUP98,* with the putative RNA helicase case, *DDX10.* Blood. 1997;89:3936-3944.
8. Raza-Egilmez SZ, Jani-Sait SN, Grossi M, Higgins MJ, Shows TB, Aplan PD. *NUP98-HOXD13* gene fusion in therapy-related acute myelogenous leukemia. Cancer Res. 1998;58:4269-4273.
9. Ahuja HG, Felix CA, Aplan PD. The t(11;20)(p15; q11) chromosomal translocation associated with therapy-related myelodysplastic syndrome results in an *NUP98-TOP1* fusion. Blood. 1999;94: 3258-3261.
10. Hussey DJ, Nicola M, Moore S, Peters GB, Dobrovic A. The (4;11)(q21;p15) translocation fuses the *NUP98* and *RAP1GDS1* genes and is recurrent in T-cell acute lymphocytic leukemia. Blood. 1999;94:2072-2079.
11. Nakamura T, Yamazaki Y, Hatano Y, Miura I. NUP98 is fused to PMX1 homeobox gene in human acute myelogenous leukemia with chromosome translocation t(1;11)(q23;p15). Blood. 1999; 94:741-747.
12. Ahuja HG, Hong J, Aplan PD, Tcheurekdjian L, Forman SJ, Slovak ML. t(9;11)(p22;p15) in acute myeloid leukemia results in a fusion between *NUP98* and the gene encoding transcriptional coactivators p52 and p75-lens epithelium-derived growth factor. Cancer Res. 2000;60:6227-6229.
13. Arai Y, Kyo T, Miwa H, et al. Heterogenous fusion transcripts involving the NUP98 gene and HOXD13 gene activation in a case of acute myeloid leukemia with the t(2;11)(q31;p15) translocation. Leukemia. 2000;14:1621-1629.
14. Ahuja HG, Felix CA, Aplan PD. Potential role for DNA topoisomerase II poisons in the generation of t(11;20)(p15;q11) translocations. Genes Chromosomes Cancer. 2000;29:90-105.
15. Jaju RJ, Boultwood J, Oliver FJ, et al. Molecular cytogenetic delineation of the critical deleted region in the 5q- syndrome. Genes Chromosomes Cancer. 1998;22:251-256.
16. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning. A Laboratory Manual. New York, NY: Cold Spring Harbor Laboratory Press; 1989.
17. Huang N, vom Baur E, Garnier JM, et al. Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J. 1998;17:3398-3412.
18. Radu A, Moore MS, Blobel G. The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. Cell. 1995;81:215-222.
19. Aasland R, Gibson TJ, Stewart AF. The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci. 1995; 20:56-59.
20. Rea S, Eisenhaber F, O'Carroll D, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. 2000;406:593-599.